DOCUMENT RESUME

ED 346 024 SO 022 353

AUTHOR La Pierre, Sharon D.
TITLE Mastery-Level Measurement: An Alternative Approach to Norm-Referenced Intelligence Testing Methods.
PUB DATE 92
NOTE 8p.; This paper was published in: "New Waves of Research in Art Education," Mary Stokrocki (Ed.), (Computer disk publication), Reston, VA: National Art Education Association, Seminar for Research in Art Education.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Art Education; *Criterion Referenced Tests; Educational Research; Intelligence Quotient; Intelligence Tests; *Mastery Tests; *Measurement Techniques; *Norm Referenced Tests; Test Construction; Test Interpretation

ABSTRACT

The measurement of intelligence (I.Q. testing method) has been based on test item construction methods that set norms for mental levels. Individual performance levels are determined by the distribution of scores based on the total group or mean score. The problem with this method of assessment for research purposes is that idiosyncratic characteristics are difficult to determine because the collective attributes or group characteristics become the norm. Individual scores fluctuate or vary due to such factors as cultural differences, test item exposure, training levels, genetics and predisposition and awareness levels, besides day-to-day events. As an alternative approach to this method, this paper explores a framework for applying mastery-level or criterion-referenced testing methods to the measurement of intelligence. Several examples were used to illustrate the potential of this method. Research done by La Pierre (1988) on spatial intelligence was used as one example. Other examples were given from research done by Gardner and Hatch (1989). The arts were used in this paper as a rich resource in regard to research on mastery-level methodology because the mastery of techniques, principles, style development as related to individuality, are key components embedded within art study. Thus the testing of mental abilities becomes the enhancement of potential through education rather than a limited concept of achievement based on the use of norms. A list of eight references is included. (DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED346024

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
SHARON D. LAPIERRE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Mastery-Level Measurement: An Alternative Approach to Norm-Referenced Intelligence Testing Methods

Sharon D. La Pierre

This paper was published in the following:

La Pierre, S. D. (1992). In Mary Stokrocki (Ed.), New waves of research in art education (computer disk publication). Reston, VA: National Art Education Association, Seminar for Research in Art Education.

SO 022 353

BEST COPY AVAILABLE

Mastery-Level Measurement: An Alternative Approach to Norm-Referenced Intelligence Testing Methods

Sharon D. La Pierre

The use of norm-referenced testing method for the evaluation of intelligence does not allow for a true appraisal of an individual's performance level due to idiosyncratic considerations. With this method of testing, the individual's raw score is related to the overall mean score. As an alternative approach to this method, this paper explored a framework for applying mastery-level or criterion-referenced testing methods to the measurement of intelligence. Several examples were used to illustrate the potential of this method. Research done by the writer on spatial intelligence was used as one example (La Pierre, 1988). Other examples were given from research done by Gardner and Hatch (1989). The arts were used in this paper as a rich resource in regard to research on mastery-level methodology because the mastery of techniques, principles, style development as related to individuality, are key components embedded within art study. Thus the testing of mental abilities becomes the enhancement of potential through education rather than a limited concept of achievement based on the use of norms.

The measurement of intelligence (I.Q. testing method) has been based on test item construction methods that set norms for mental levels. Individual performance levels are determined by the distribution of scores based on the total group or mean score. The problem with this method of assessment for research purposes is that idiosyncratic characteristics are difficult to determine because the collective attributes or group characteristics become the norm. Individual scores fluctuate or vary due to such factors as cultural differences, test item exposure, training levels, genetics and predisposition and awareness levels, besides day-to-day situational events. As a matter of fact, norm-referenced tests such as many intelligence measurements depend on the existence of difference among subjects' true scores. "When we use norm-referenced measures, we want to know how far a student's score deviates from the group mean" (Livingston, 1972). The reason for the deviation is not a concern. An individual's level of mental performance is just classified and compared to his or her peer group based on age considerations.

The historical aspect of measuring intelligence "as a single quantity" has existed as an expression of and search for objective knowledge (Gould, 1981). Gould (1981) criticized "the myth that science itself is an objective enterprise, done properly only when scientists can shuck the constraints of their culture and view the world as it really is" (21). Science, the area of research being one component of this discipline, is embedded in a socially prescribed enterprize. Humankind has attempted to measure the qualitative levels of thinking for decades by developing theories and the ultimate statistical analyses to determine the essential properties and hierarchy of thought.

Cole (1985) does not believe that the mind can be "...assessed independently of culturally organized experience" (219). I.Q. testing is at present a random sample of activities that are required of school learning and not necessarily representative of life's experiences across all cultures. Hypothetical items are not the same as real operations and levels of progress are researched as conditions of mental development, exclusive of environment, experience, background, and specific domain competence (contexts and expertise) that may not be subject to the laws of hypothetical knowledge. According to Gardner and Hatch (1989), "Most definitions of intelligence focus on the capacities that are important for success in school. Problem solving is recognized as a critical component, but the ability to fashion a product—to write a symphony, execute a painting, stage a play, build up and manage an organization, carry out an experiment--is not included, presumably because the aforementioned capacities cannot be probed adequately

Copyright © 1992 by Sharon D. La Pierre
4793 Briar Ridge Trail
Boulder, Colorado 80301
All Rights Reserved.

in short-answer tests" (5).

If research on the nature of intelligence continues to utilize norm-referenced testing method as a primary way of assessment, no avenue will exist for the exploration of individual thinking qualities based on degrees of mastery that have the potential of expanding to other levels of achievement. What "is" and what "could be" are two different concepts to be considered in the process of researching intelligence factors. An I.Q. test score is not the utimate statement of fact in regard to an individual's ability to process thoughts on an intelligent level. It must be recognized that norm-referenced I.Q. testing is only one method used to research and assess the abilities of human beings.

The purpose of this paper was to explore the use of criterion-referenced testing method (or mastery levels of performance) as a possible research direction for the measurement of intelligence factors. The focus of this discussion centered on the possibility of "setting" standards as opposed to imposing "a particular" standard. The assumption that was made was that norming common characteristics does not allow individual characteristics to surface. Eisner (1985) made a similar point when he stated the following:

> To cast a net into the sea that is unintentionally designed to let most of the fish get away, and then to conclude from those that are caught of what the variety of fish in the sea consists is, at the very least, a sampling error of the first order. Then to describe the fish that are caught in terms of their length and weight is to reduce radically what we can know about the qualitative features of the ones that have been caught, not to mention the features of those that the net failed to catch in the first place. (152-153)

The focus of the arts has much to offer in the way of criterion-referenced or mastery level research methodology because the mastery of techniques, principles, style development as related to individuality, are key components embedded within the study of art. The assumption made in this paper is that mastery or the development of competence levels allows for the unfoldment of potential abilities, not allowing the individual expresser of acquired knowledge to be embodied or obscured within the group score.

Mastery-Level or Criterion-Referenced Measurement

The evaluation of individuals' levels of proficiency characterizes mastery-level test method. It is also refered to as criterion-referenced measurement because the objective of such test methods is to see how far a true score deviates from a fixed standard or criterion score (Livington, 1972). Therefore it is necessary to establish some sort of standard for performance. The judgement for this kind of standard is based on a reasonable and subjective approach according to Millman (1973). What that means is that levels, states, or categories (such as mastery level, partial-mastery level, or non-mastery level) may be established. These designated categories are established by the setting of cutoff scores that accommodate a statistical framework based on criterion scores or fixed standards. The individual examinee's score is then compared to the categories' cutoff scores for proper placement.

It must be realized that an examinee's true mastery state is a probabilistic situation and that misclassification can occur. However, this is also true of any other evaluation procedures as well. One major limitation of criterion-referenced measures is getting experts to agree as to the degrees and levels of mastery. However, it must be remembered that the arts have a long history of functioning with the use of experts as a valid way of assessing performance; this practice is usually seen in the form of jurors or judges and critiques. In research terms this is what is refered to as rater scoring or interrater reliability, the use of independent observers to score a test in order to establish dependable scoring or evaluative results.

Examples of the Use of Criterion-Referenced Test Method

Examples that used the mastery-level evaluation process to measure various aspects of intelligence are described in this paper. The first example is a study conducted by the writer to measure the intellectual process of spatial reasoning and the summary that follows is condensed to include mainly the analysis section that dealt with the use of criterion-referenced cutoff scores after the study was run (La Pierre, 1988). The use of criterions as a measurement should be supported by an extensive theoretical

framework to justify the classifications as reasonable standards. This particular study on spatial reasoning had an extensive conceptual base but is not elaborated on in this paper.

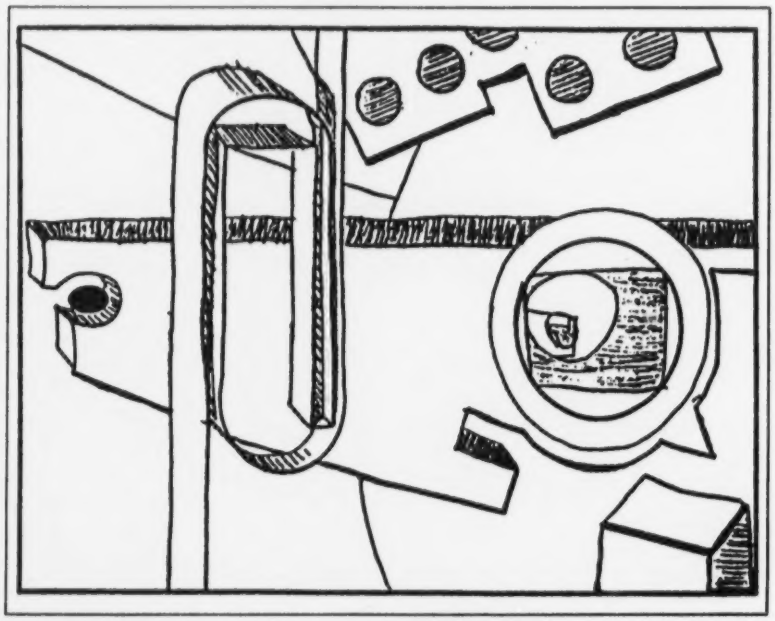
The subjects of this first example totaled 109 students from Colorado State University, Ft. Collins, Colorado. The average age was 27.3 years, ranging from 18 to 58 years. There were 72 females and 37 males in the total group. The subjects were divided into two distinct groups. Fifty-five (55) subjects made up the art group, consisting of advanced fine arts students from painting and printmaking courses (what would be called "applied" art). Fifty-four (54) subjects made up the English group, consisting of advanced English students from literature, essay, and semantics courses. There were 37 females and 18 males in the art group and 35 females and 19 males in the English group. Five subjects declared another language as their native tongue, four from the English group and one from the art group. Seventeen (17) subjects had had some technical training, ten from the English group and seven from the art group. Fourteen (14) of the subjects from the English group had had art training, either in high school or college (one had a BFA degree). The average age of the art group was 26.2, ranging from 18 to 58 years, and the average age of the English group was 28.4, ranging from 19 to 50 years.

Although the sample group from the arts was composed of "applied" visual art students, the results of this study may or may not have applied to students taking classes in art history, critique method, or aesthetics. It must be recognized that different areas of the visual arts consist of intelligence in the verbal as well as the non-verbal or figural mode. Spatial reasoning as defined in this study is one aspect of the thinking process.

The following is a summation of this study and how it relates to the purpose of this paper.

FIRST EXAMPLE: The purpose of this study was to investigate the process of spatial reasoning in the adult through the analysis of the level of manipulation of abstract mental imagery by assessing a "doodle" activity. The doodle activity was based on a conceptual framework regarding an individual, either child or adult, as reproducing markings naturally as a result of mental imagery that is usual to his or her own understanding or keenness of mental perception in terms of spatial organization. This organization of space is regarded as an intelligent act used to think and to solve problems without the use of the discursive mode, but rather, based on figural structure (La Pierre, 1988, p. 1).

Two raters scored the doodle activity using a Spatial Measurement Scale that was developed as a result of a conceptual framework based on design elements extrapolated from the visual arts. The study examined the applicability of the Measurement Scale in determining an individual's level of spatial ability. The results of this study indicated that spatial reasoning was highly related to figural structures that allowed for the mental manipulation of abstract mental imagery as a reasoning faculty. The results of this study also indicated that significant interrater reliability (.76) existed on the Measurement Scale and that the groups tested were significantly different, resulting in the art subjects scoring higher due to mastery level (La Pierre, 1988, p. 1). Refer to Figures 1 and 2 for an example of the comparison made between the doodles of the two tested groups.

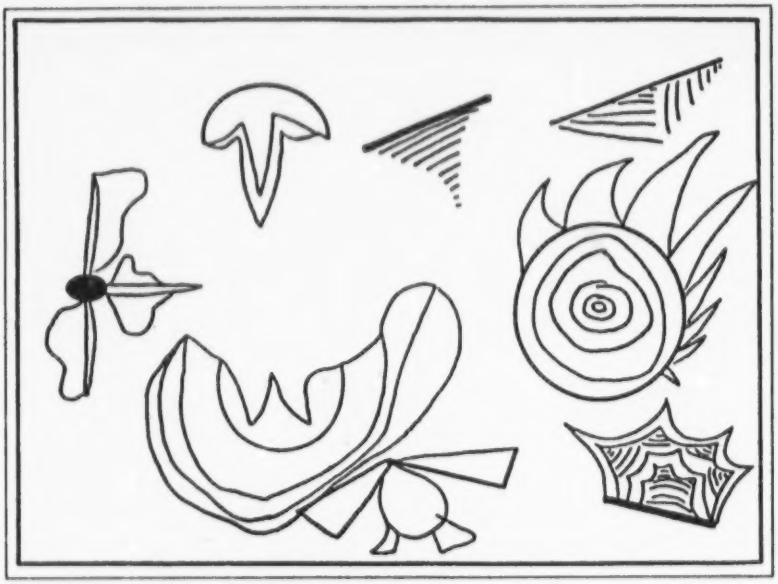


GROUP ONE

Figure 1

An example of the comparison between the doodles of the two tested groups. This particular item is from the art group.

BEST COPY AVAILABLE



GROUP TWO

Figure 2

An example of the comparison between the doodles of the two tested groups. This particular item is from the non-art group.

Part of the analysis in this study included the setting of cutoff scores. Three levels of spatial functioning, states, or categories were established. They included *mastery-level* (advanced proficiency), *partial-mastery level* (intermediate proficiency), and *non-mastery level* (remedial proficiency). The Measurement Scale had 29 items to score. The two raters' scores were arithmetically averaged for each item. Each item was weighted because the importance of the items varied; not all items were of equal importance. The weighted scores for each item were multiplied by the averaged raters' scores and totaled for an overall score. This cumulative total score was used to place the subjects in one of the three levels of mastery. The art group (based on ranked total scores) was considered to be the master group and was used to set the standard performance level for another group of examinees, as well as for any other subsequent groups to be tested. It was determined that an examinee must have scored at least as well as the bottom 10% of the art group scores to be considered partial-mastery level. A mastery level classification required a score of at least 85% or above of the art group. Frequency distributions were calculated to determine the percentage of the examinees in this study, other than the art group, that fell within each group classification. Refer to Figure 3 for a visual clarification of level percentages and group classifications as set by criterion-referenced scores (La Pierre, 1988, p. 23).

This study used other statistical procedures to verify the mastery level of the art group. Factor analysis, analysis of variance, and Pearson Product-Moment Correlation confirmed the results of the criterion-referenced procedure.

According to Gardner and Hatch (1989), "...separate psychological processes appear to be involved in dealing with linguistic, numerical, pictorial, gestural, and other kinds of symbolic systems...Individuals may be precocious with one form of symbol use, without any necessary carryover to other forms" (5). If this is indeed the case, then the use of measuring intelligence based on specific kinds of domains or psychological functions (such as spatial reasoning) can best be assessed through the use of proficiency levels to establish how well the individual has mastered particular thinking processes at certain times.

SECOND EXAMPLE: According to Gardner and

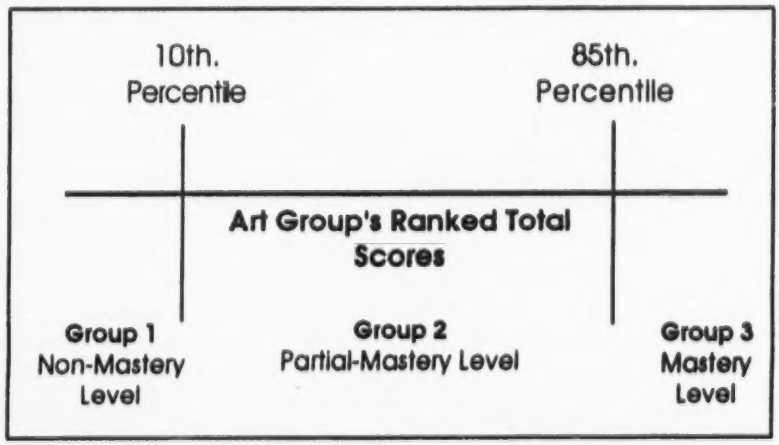


Figure 3

Mastery levels based on the art group's ranked total scores.

Hatch (1989), between 1987 and 1989 field tests were used in classrooms to investigate the hypothesis that intelligences are independent of each other. The purpose of these various studies was to varify Gardner's theory of Multiple Intelligences (MI) as detailed in *Frames of Mind* (Gardner, 1983). Assessment of each kind of intelligence, to identify their strengths and weaknesses, were developed—to include logical-mathematical, linguistic, musical, spatial, bodily-kinesthetic, interpersonal, and intrapersonal. One conclusion reached included the following (Gardner & Hatch, 1989):

> In our work, it rapidly became clear that meaningful assessment of an intelligence was not possible if students had little or no experience with a particular subject matter or type of material. For example, our examination of bodily-kinesthetic abilities in a movement assessment for preschoolers was confounded by the fact that some 4-year-olds had already been to ballet classes, whereas others had never been asked to move their bodies expressively or in rhythm. This recognition reinforced the notion that bodily-kinesthetic intelligence cannot be assessed outside of a specific medium or without reference to a history of prior experiences. (6)

It was concluded from various observations that in order to evaluate intelligence properly, activities must be given to students to increase the students' chances of experiencing and engaging for the purposes of achieving some success. Projects which encompassed various schools such as Arts PROPEL and Project Spectrum were developed for this purpose. Criteria were developed for assessment purposes. Because the programs were not in final form, the evaluation and results reported were still in preliminary form. What is of importance in the writing of this article, is the fact that such assessment is based on criteria. One specific study from these projects is as follows (Gardner and Hatch, 1989):

> Eight kindergartners (four boys and four girls) and seven first graders (five girls and two boys) were assessed on the seven activities of the Modified Spectrum Field Inventory (MSPFI). This inventory, based on the activities developed for the yearlong Spectrum assessments of preschoolers, consists of activities in the areas of language (storyboard), numbers and logic (bus game), mechanics (assembly), art (drawing), music (xylophone games), social analysis (classroom model), and movement (creative movement). These assessments were administered in two 1-hour sessions. Each activity was videotaped, and children were scored by two independent observers. Spearman rank-order correlations between the scores of the two observers ranged from .88 (language) to .97 (art) and demonstrated the interrater reliability of these scores.
>
> ...an examination of children's ranks on each of the activities revealed a more complex picture...considered in this way, children did exhibit relative strengths and weaknesses across the seven activities. (8)

The goal of this work was to detect specific human strengths (or intelligences) and to use these strengths as the basis for engagement and learning (Gardner & Hatch, 1989).

Summary

In conclusion, while lecturing on mastery in relationship to children, Lowenfeld (1985) made the following observation:

> Let us always remember that they [meaning children] are individuals. When the child scribbles—all children scribble, but by the very way children scribble differently we can see much of their behavior, if they are allowed to scribble freely.....[The child] discovers that he can control his motions and that there is a connection between the line he produces on the paper and the movements he makes with his arm. This is a great discovery for him, because at that time, unconsciously, he discovers for the first time that he can master his movements. Now, he gets lots of thrill out of it; and, of course, whenever we would like to show ourselves that we are able to master a situation, what do we do? We repeat it, again and again. You know how often you can tell a story to your children without their getting tired. Again. Again. Again. You know this is quite understandable for children, because of the desire for self-assurance, mastery of a situation. (23)

Perhaps we in the arts should be encouraging the development of "virtuosity" and not just general enrichment or overall instructional methods that do not allow individual mental processing to develop as unique strengths. Historically the arts have been the center for the development of disciplined activity. This kind of learning has incorporated the use of criteria to master various levels of development---or a progressive learning process. Perhaps these levels of mastery exemplify the expression of specific intelligences where degrees or levels of mastery depend on engagment and exposure to such activities. We as educators will be better able to promote efforts, potential abilities and various stages or levels of mental growth that are not yet differentiated (Lowenfeld, 1985). Thus, the testing of mental abilities becomes the enhancement of potential through education rather than a limited concept of achievement based on the use of norms.

References

Brittain, W. Lambert (Ed.). (1985). Viktor Lowenfeld speaks on art and creativity. Reston, Virginia: National Art Education Association.

Cole, M. (1985). Mind as a cultural achievement: Implications for IQ testing. In Elliot Eisner & Kenneth J. Rehage (Eds.), Learning and teaching the ways of knowing: Eighty-fourth yearbook of the National Society for the Study of Education (pp. 218-249). Chicago, Illinois: The University of Chicago Press.

Eisner, E. W. (1985). The art of educational evaluation: A personal view. Philadelphia: The Falmer Press.

Gardner, H., & Hatch, T. (1989). Multiple intelligences go to school: Educational implications of the theory of multiple intelligences. Educational Researcher, 18 (8), 4-10.

Gould, S. J. (1981). The mismeasure of man. New York: W. W. Norton & Company.

La Pierre, S. D. (1988). Spatial reasoning and its measurement. In Robert A. Fellenz (Ed.), Spatial reasoning and adults (pp. 1-31). Monographs of the Kellogg Center for Adult Learning Research, Montana State University, Bozeman, Montana, 2.

Livingston, S. A. (1972). Criterion-referenced applications of classical test theory. Journal of Educational Measurement, 9 (1), 13-26.

Millman, J. (1973). Passing scores and test lengths for domain-referenced measures. Review of Educational Research, 43 (2), 205-216.